

# In-vehicle coupon recommendation data set

...

IronHack Data Analytics Final Project

11/03/2022

by Nina Thiessen

Imagine you're driving your car and receive an offer for a coupon for a 20% discount at a local bar or restaurant.

## Do you accept the coupon offer?

...the answer is *“It depends”*

## Project goals

1. Develop a classifier that accurately predicts whether or not a driver would accept the coupon in various scenarios
2. Identify the factors that have the biggest influence on this decision

(“Accept” means they intend to use the coupon before it expires)

# Data used for modeling

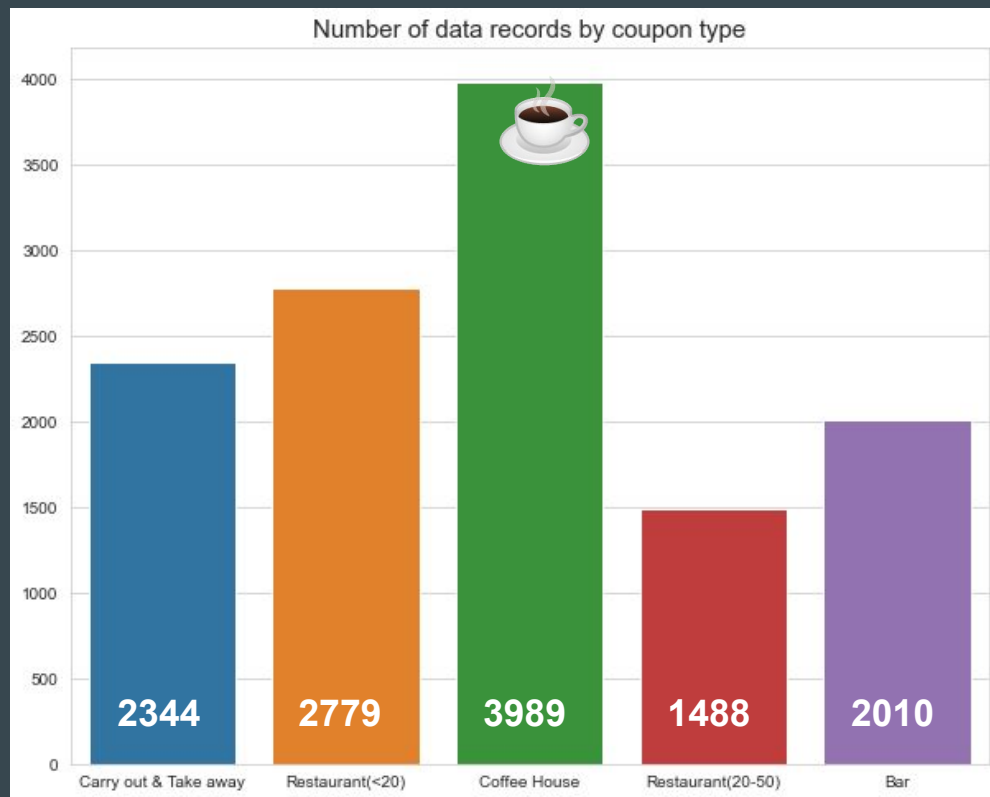
21 categorical features

12,610 unique records

5 distinct subsets based on coupon type:

1. Take-out
2. Cheap Restaurant (<\$20)
3. Coffee House
4. Expensive Restaurant (\$20-50)
5. Bar

The data was collected via survey to assess the feasibility of an in-vehicle coupon recommender system.



Wang, Tong, Cynthia Rudin, Finale Doshi-Velez, Yimin Liu, Erica Klampfl, and Perry MacNeille. 'A bayesian framework for learning rule sets for interpretable classification.' The Journal of Machine Learning Research 18, no. 1 (2017): 2357-2393.

# 9 Contextual attributes

## **Coupon type:**

Bar, Coffee House, Take-out, Cheap Restaurant (<\$20), Expensive Restaurant (\$20-50)

**Coupon expiration:** 2 hours, 1 day

## **Coupon destination**

**Distance :** 5-14, 15-24, 25+ (minutes drive)

**Direction:** same, opposite (as current destination)

**Destination:** Home, Work, No Urgent Place

**Passenger:** Alone, Friend(s), Kid(s), Partner

**Weather & Temperature:** Snowy & cold, Rainy & cool, Sunny & cold/cool/hot

**Time:** 7am, 10am, 2pm, 6pm, 10pm

# 12 Driver attributes

## Personal demographics

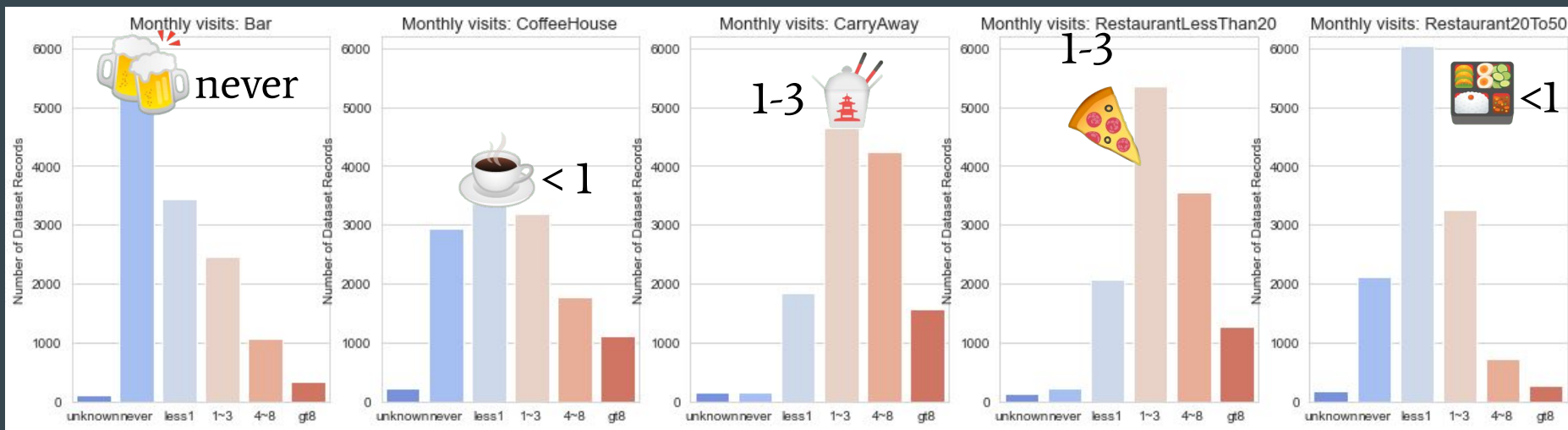
Gender, Age, Marital status, Has children?, Education level, Occupation, Income

## Personal preferences

Monthly frequency of visits for each of the 5 coupon types. i.e.

- *How many times per month do you normally get take-out?*
- *How many times per month do you normally eat at a restaurant with average expense less than \$20 per person?*
- *etc...*

# Replace missing values in monthly visit preferences with mode



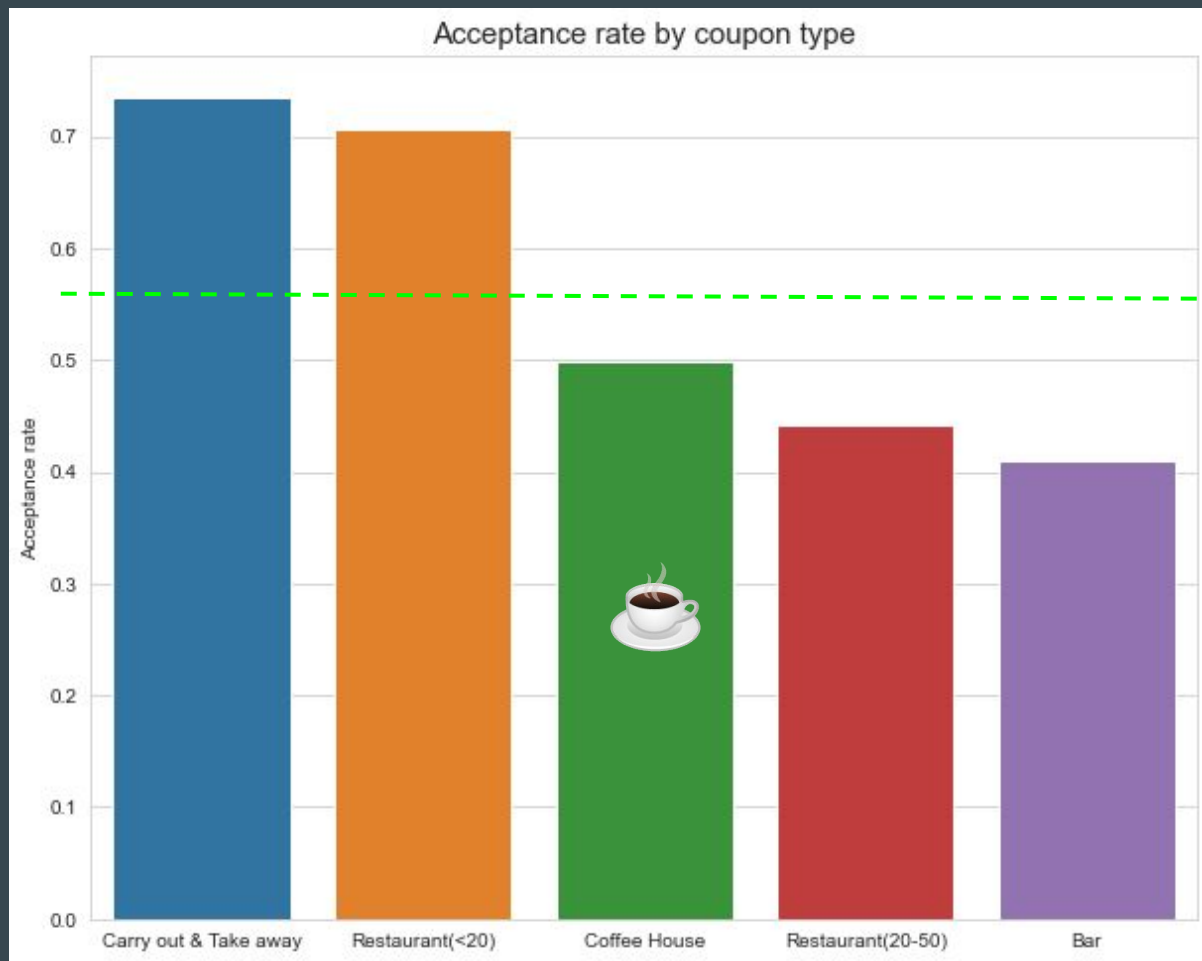
~5% of the records had one or more missing values in the five monthly visit frequency features  
→ These missing values in were replaced with the most common response

In addition, in case these values were not missing at random, five new ‘frequency unknown’ indicator variables were created.

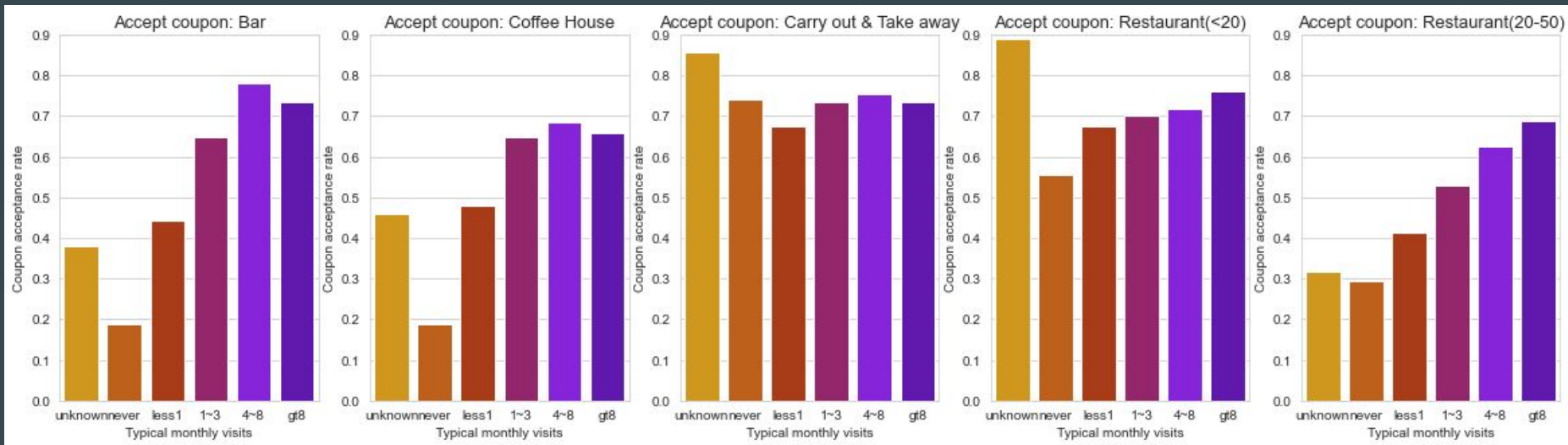
## Overall coupon acceptance rate is 57%

However the acceptance rate varied substantially within each of the 5 data subsets.

In order to ensure more accurate modeling SMOTE upsampling was used to balance the data

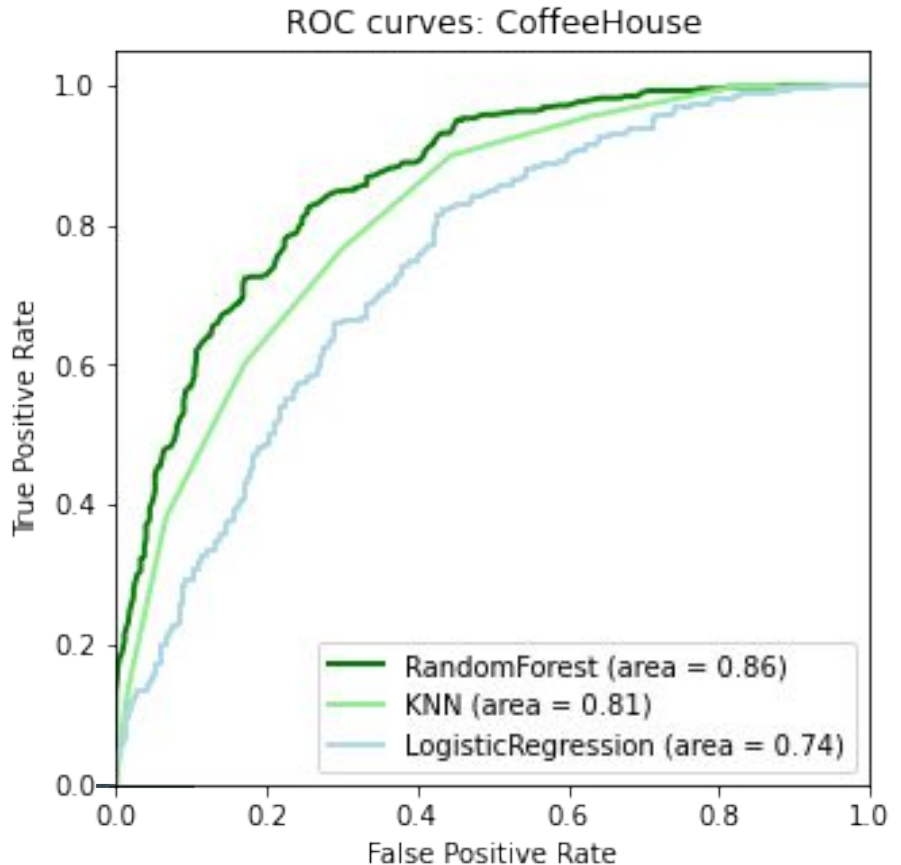


# Acceptance rate vs Monthly visit frequency





# Selecting the Classifier with highest Area Under ROC Curve








# ROC AUC for RandomForestClassifier optimized with GridSearchCV

Train/test split 80/20

5-fold Cross Validation






Hyper parameter tuning:

- max\_features (2-15)
- min\_samples\_split (2-5)
- n\_estimators (100, 150, 200)

Coupon type	CV mean	CV std	Test set
 <b>Restaurant (\$20-50)</b>	<b>0.77</b> 1253	0.016	<b>0.77</b> 4452
 <b>Coffee House</b>	<b>0.84</b> 8320	0.013	<b>0.86</b> 1093
 <b>Bar</b>	<b>0.89</b> 5877	0.013	<b>0.91</b> 3444
 <b>Take-out</b>	<b>0.91</b> 9796	0.012	<b>0.92</b> 2260
 <b>Restaurant (&lt;\$20)</b>	<b>0.92</b> 3142	0.010	<b>0.94</b> 8076






# Feature Importance: Personal preferences matter

## Bar

	feature_name	importance
0	Bar 	0.133
1	CoffeeHouse 	0.047
2	Restaurant20To50 	0.042
3	CarryAway 	0.042
4	RestaurantLessThan20 	0.041
5	minsToCouponDest	0.031
6	has_children	0.030
7	gender_Male	0.021
8	maritalStatus_Single	0.020
9	maritalStatus_Married partner	0.019
Result: Bar		



## Coffee House

	feature_name	importance
0	CoffeeHouse 	0.139
1	time	0.048
2	education	0.046
3	minsToCouponDest	0.045
4	Bar 	0.043
5	RestaurantLessThan20 	0.041
6	CarryAway 	0.040
7	Restaurant20To50 	0.039
8	expiration_2h	0.038
9	temperature	0.026
Result: CoffeeHouse		

# Feature Importance: Personal preferences matter...sometimes

## Restaurant (\$20-50)

	feature_name	importance
0	income	0.076
1	CoffeeHouse	0.061
2	Restaurant20To50	0.061
3	education	0.054
4	CarryAway	0.051
5	RestaurantLessThan20	0.050
6	Bar	0.050
7	time	0.046
8	expiration_2h	0.041
9	temperature	0.034

Result: Restaurant20To50

## Restaurant (<\$20)

	feature_name	importance
0	income	0.061
1	age	0.058
2	minsToCouponDest	0.056
3	CoffeeHouse	0.054
4	education	0.051
5	RestaurantLessThan20	0.050
6	CarryAway	0.049
7	Bar	0.047
8	Restaurant20To50	0.046
9	destination_No Urgent Place	0.043

Result: RestaurantLessThan20

## Carry away/ Take-out

	feature_name	importance
0	income	0.070
1	education	0.062
2	CoffeeHouse	0.062
3	age	0.061
4	RestaurantLessThan20	0.056
5	time	0.056
6	CarryAway	0.055
7	Restaurant20To50	0.053
8	Bar	0.051
9	minsToCouponDest	0.047

Result: CarryAway

# Thank you

Questions?

